REVIEW OF ALASKA POLLOCK ASSESSMENT

T.Kevin Stokes
CEFAS Lowestoft Laboratory, Pakefield Rd, Lowestoft, NR33 0HT, United Kingdom

SCOPE OF REVIEW AND HOW APPROACHED

The purpose of the review was to consider the assessment methods used for Bering Sea pollock. The emphasis has been on identification of deficiencies and possible new efforts that might be undertaken.

In order to carry out the review, various documents were made available in advance (either electronically or by post), and during, meetings at the AFSC in Seattle over the period 17-21 January 2000. During the time spent in Seattle, a large number of presentations were made by, and discussions held with, AFSC staff and expert colleagues from the University of Washington. A considerable amount of ground was covered on all aspects of the fishery, biology, assessment methods, management context and current advisory framework.

As a single reviewer, working closely with the senior assessment scientist (Jim Ianelli), the procedures and various discussions were kept informal. I believe, however, that most avenues were explored exhaustively. The only meetings planned at the start of the process, that failed to take place due to lack of time, were with industry and environmental group representatives. Jim Ianelli, however, was very careful to try and present a balanced perspective, including an indication of counter-arguments to his own. I would like to thank Jim Ianelli, all AFSC staff and University of Washington staff, for their time, openness and help. Of course, I take full responsibility for all comments and views expressed below.

This report forms part of the contract agreed with the University of Miami. My hope, as reviewer, is that the discussions in Seattle were the most important part of the process, and that those discussions, rather than any written words, will have led to further thinking and work being undertaken. The written report constitutes a formal record of what was covered in Seattle, but does not add anything substantive to those issues already discussed at length.

Kevin Stokes
Lowestoft
February 2000

FISHERY DATA

(Discussions took place with Jim Ianelli, Dan Kimura and Sarah Gaichas of the AFSC, Don Gundersson and Vivian Haist of the University of Washington)

Recent/future data

The existing observer program provides excellent coverage of the offshore fleet. It is clear that considerable care and attention has been given to the program and that planning for the future is well in hand. The work of Kimura (1989) has laid the basis for sample sizes to achieve given coefficients of variation (CVs) on the catch-age re-constructions. Generally, a target CV of five per cent requires about five hundred otoliths *per stratum* to be read, where a *stratum* in this case is defined by sex and region – there are six for Eastern Bering Sea (EBS) pollock. Although a very large number of fish have been sampled for length, and otoliths collected, generally only 2000-2500 otoliths have been read *per stratum*, resulting in achieved CVs of approaching ten per cent. Although slightly "off target", this is still excellent by most catch-age data standards and provides a good basis for the subsequent assessments.

There has been considerable over-collection of otoliths, and slightly inadequate otolith reading. The new collection regime should streamline the procedures - but the main constraint will still belimited capacity for onshore otolith reading. With fishery practices likely to change due to implementation of the American Fisheries Act (AFA) and Reasonable and Prudent (management) Alternatives (RPAs) for Steller sea lions, it may be that more spatially explicit models, fit to shorter time-scales will be required. If this were to be the case, more detailed *strata*-specific information would be required from the fishery, and the sampling effort would need to be increased

It was noted that under the new management regime (AFA), a greater proportion of inshore boats with no observers would be participating in the fishery. Furthermore, because of the complicated restricted areas to protect Steller sea lions, these inshore boats would be forced further offshore than their traditional grounds, with the possibility of off-loading catches at sea. Questions of enforcement and data quality arise. Issues such as these are complicated and it is difficult to second-guess, let alone forecast, their effect on the pattern of the fishery. It is clear, however, that the assessment in 2000, and forecasts for ABCs, may be sensitive to changes in the fishery.

Early data

In the early years, fishing was primarily by Japanese boats for the surimi market. The landings and catch-age data are relatively poor, and there seems to be some question as to exactly how the numbers were derived. These data should be re-examined as to their origins and, if necessary, re-computed (There is some suggestion that CPUE data from standard Japanese trawlers during the 1960s might be available; this should also be investigated). Given their unreliability, the question arises as to whether these early data should be used in the assessment. In this regard, there is a trade-off between the overall reliability of the data *versus* the indication that the EBS pollock biomass was probably at low levels (prior to any intensive fishing) in the early 1960s. But the apparently accurate estimates of biomass in the early 1960s could be misleading. Given the apparent regime shift of 1978, poor early-year catch-age data,

and the low length sampling in the 1979 EIT (Echo-Integration Trawl surveys; see below), it may be sensible to carry out the assessment from 1980 only. This needs to be investigated.

Other data issues

Russian data from the Northern Bering Sea (NBS) are not available but the (international) fishery in the area takes of the order of 0.5 Mt *per annum*. As the NBS and EBS are apparently continuous (BIOLOGY below), this is potentially a major problem (see also MANAGEMENT below).

The level of discarding in the pollock fisheries is historically relatively low (compared to many fisheries world wide) and is declining. So long as discarding remains at this low level, it is not considered likely to be a problem in the estimation of relevant quantities of interest for the setting of an ABC or for longer term forecasting.

SURVEYS

(Discussions took place with Jim Ianelli, Neil Williamson, Taina Honkalehto, Dave Somerton)

BTS

The bottom trawl survey is neither random nor stratified random, but consists of fixed stations on a well-worn grid, fished with 30 minute (on the ground) tows. The gear is low headline, designed for catching flatfish, but pollock are semi-pelagic in habit. *A priori* the indices should not therefore be too good for pollock; nevertheless, from the assessment diagnostics, they appear to be of good quality. The survey apparently "picks up" 1 year-old fish quite well, but not 2 year-olds and 3 year-olds. The exception to this pattern is that the survey also picked up many 2 year-olds in 1998 and 3 year-olds in 1999 – surprisingly, therefore, the survey has had little apparent effect on recruitment (as 1 year-olds) estimation for 97. The sampling levels for ageing material appear to be appropriate.

ECHO-INTEGRATION TRAWL SURVEYS (EIT)

In general, the EIT has not extended into the NBS[1]. Given the lack of fishery data from the NBS, the EIT potentially has an important role to play in the assessment if it can be extended. Inter-calibration between US and Russian (older) boats has taken place. The Russians, however, now use a different boat. There is clearly a need for better EIT (, survey, and fishery) coverage in the NBS, and the assessment needs to be of the NBS/EBS combined. From discussions, it is clear that AFSC staff are trying hard to coordinate with their Russian counterparts to achieve this; this needs to be encouraged.

[1] The biomass in the NBS is of the order of 0.5 to 1.0 Mt, plus whatever is "right on the bottom". The exploitation rate (~0.5 Mt *per annum*) in the NBS is therefore relatively high compared with the EBS.

The trawl (length) sampling in the first year of the EIT (1979) was low and may be a factor in the poor model fit to EIT data in that year. The weightings for age-composition used in the assessment (MODELLING below) already take account of the different sampling levels in the fishery, BTS and EIT for most years – but for 1979 the EIT receives the same weight as in subsequent years. The 1979 EIT datum could be down-weighted further, or simply deleted (this is the simplest option).

BIOLOGY

(Discussions[2] took place with Jim Ianelli, Kevin Bailey)

Pollock in the North Pacific are extensively distributed with a diversity of niches. Numerous small populations exist in isolated regions but the large EBS stock is central in distribution. There is considerable phenotypic and genetic evidence that the EBS is linked to the Gulf of Alaska and NBS populations, but that it is otherwise separate. Although the biology is fascinating, the clear issue of importance for EBS management, is that of Russian catches and data for the NBS. This clearly needs to be addressed and clarified before there can be full confidence in the assessment and management advice.

MODELING

(Discussions took place with Jim Ianelli, Vivian Haist)

Assessments

Whilst assessments are intrinsically interesting, and many scientists want them to provide the "best" biological insight, the most important thing to achieve through their use is robust advice. That is, the sensitivity of key management quantities (in this instance, $Bcur/B_{MSY}$; F_{MSY} etc.) to model structure, data and implementation choices needs to be checked. However, there is also a need to determine a "best" model that can be used for a period of time with minimal changes – both to ensure consistency and stability of advice, and to build credibility (in the science and in the decision-making). Work undertaken in the pollock assessment is clearly of the highest standard in both respects. Some changes in details of implementation have been made in 1999, and need to be fully documented. (For example, in the 1999 assessment, year-specific effective sample-sizes for age composition data for the surveys and fishery data were added to the model but not documented. This primarily affected the fidelity to fitting the early fishery data periods where the age composition information is considered less reliable.) The main assessment model, however, is essentially unchanged from last year.

[2] The FOCI programme was not covered during these discussions.

Quasi-Bayesian approach – the use of approximate methods (using AD-Model Builder (ADM)) versus full Bayesian integration (using the Markov chain Monte Carlo (MCMC) method). There are major advantages with using a Bayesian approach (the calculation of intuitively interpretable probabilities, a natural feed-in to decision analyses, the use of prior information from a variety of sources etc.) as opposed to more traditional fishery assessment models. Computation time, however, is a problem, especially for full sensitivity testing. The approach adopted for the EBS pollock assessment is a good one – use a fast approximation to the Bayesian analysis, still using prior information, to investigate alternative model structures and sensitivity to data and assumptions. In the final analysis, however, having adopted a particular model, it is then important to run the full Bayesian integration, to check fully integrated *versus* approximate posteriors. The pollock assessment documents do not indicate that the full Bayesian model was run, but during the review discussions in Seattle, comparisons were made between the fully integrated Bayesian model and the documented approximate results. For the preferred assessment model (MODEL 2, 1999) the full and approximate posteriors for F_{MSY} were different – but, perhaps fortuitously, the all- important (for ABC calculations) harmonic means of each are the same to three decimal places. Looking at the single numbers is apparently, however, insufficient; a graphical comparison is also essential. Although it appears not to make any practical difference in this particular case, in this particular year, it is important to use the approximate methods only to "screen" models and to use full Bayesian integration for final runs. Note that the Tier system (adopted by Council) has provided a risk-averse approach (under Tier 1) based on a fully Bayesian decision analysis (Amendment 56). The requirement for Tier 1 is that a "reliable" probability density function (pdf) for the F_{msy} value is available. The question naturally arises as to what is "reliable"? Is the approximate pdf reliable when it is different to the fully Bayesian pdf, even though the quantity of interest, the harmonic mean, is the same?

Sensitivity to start date -–especially of Maximum Sustainable Yield (MSY) (*via* F_{MSY} (the fishing mortality rate associated with MSY). The reliability of early year catch-age data is questionable. Together with a mooted regime shift in 1978, and poor length sampling of the 1979 EIT, this creates a potentially major problem. It is clear from runs made during the visit to the AFSC that there is a need for further exploration of the sensitivity of $Bcur/B_{MSY}$ and of F_{MSY} to the start date of the assessment, or the weightings (effective sample sizes) chosen through time for the age-composition data. This should be done before the main 2000 assessment.

Effective sample size/ weighting. There is some confusion in both the 1999 and 1998 assessment documents as to what weightings were actually used. In the 1999 document, for example, the subjectively set effective sample sizes for the age-composition estimation are stated differently on pages 12 and 76. For 1998 they were set as 100/50/25 for the fishery, BTS and EIT respectively, for the whole time series. For 1999, they were again 100/50/25 for recent years, but were set differently in the early assessment period. The effective sample sizes actually (as opposed to reported) used are apparently reasonable when compared with the effective sample size values given in the report for MODEL 2 (Table 1.9). However, in future, a clearer description and justification should be provided. Also, if the assessment is to use all data, the 1979 EIT should be down-weighted by a further reduction in the effective sample size assumed.

The steepness prior. The estimation of F_{MSY} depends on the steepness of the slope-at-origin of the R/S curve. For the preferred assessment (Model 2), using the full data series, the data/model "tries" to make the slope at origin steeper, but the prior "pulls" it back down. The apparent regime shift from 1978 led to a less productive period. If only data from 1980 onwards are used in the assessment, the steepness should therefore be expected to be less than that derived from the full data series. The posterior mode is indeed lower than that using the full time series of data, but still greater than the mode of the prior. If the assessment is run using data from 1980 onwards, but with no prior, the steepness is greater than with the prior (which is therefore seen to be influential) for the 1980 analysis only, but also greater than in the analysis with a prior for the full dataset. Clearly, the prior on steepness is very influential. As this affects the F_{MSY} posterior and harmonic mean, and eventual ABC estimation, there is a clear need to investigate further the role of the prior. This is not just a technical point – it is also a "philosophical" one. The Bayesian approach makes use of prior information – it is this that permits the eventual interpretation of outputs as probabilities. But the value lies not just in interpretation, but in the use of belief as well as data-derived knowledge. Priors need to be described, explained and justified – and ideally need to be supplied by a range of "experts" rather than a single analyst.

Over-parameterisation. The assessment attempts to estimate a large number of parameters – mostly selectivity-at-age of surveys etc. *A priori*, there is no reason to expect survey selectivities-at-age to change through time. The arguments given that the selectivity is essentially a reflection of distributional changes relative to surveyed areas is not unreasonable, but as the model is arguably over-parameterised, not estimating selectivities for the surveys is an option that should be explored. The reduction in RMSE through estimating changes in selectivity, could likely be achieved either by removing the poorly estimated 1979 EIT point, or starting the analysis from 1980 on. (As this also circumvents arguments about regime shifts, poor EIT length sampling etc, it seems a useful way forward.)

Forecasting

In 2000, and resulting from the American Fisheries Act, there will be changes in the ratio of inshore/offshore catches. Cold water in 1999, and ice cover in 2000, are likely to reduce recruitment below that assumed in 1999 (i.e. existing forecasts are likely to be optimistic). For 1999/2000 and on, the sea lion RPAs will effect the timing and distribution of the fishery – selectivity changes should therefore be expected together with a possible increase in the proportion of juveniles caught. There are some difficult problems ahead for the 2000 assessment, especially the 1999 selectivity estimation and assumptions for forecasting – both to estimate the 2001 ABC, and to estimate biomass levels over the next ten years. These factors will have to be carefully untangled and explained in the advice.

Possible presentation forms

It is too easy to criticise assessment documents. They are too short for many, too technical for others, and so on. I know from experience that there is no winning. The

basic assessment document for the EBS pollock assessment is readable and reasonably complete. I would, however, suggest the following as perhaps preferable for 2000, and if time permits.

Provide a short Council document similar to that already produced, but with the addition of a clear tabulation in the main text of the 1998, 1999 and 2000 assessment assumptions, and the "dial settings" of the final preferred model. The point is to make clear how the advice is dependent on data updates, model assumptions, or "tweaking". For this document, it is open to question whether or not extra decision-type tables might be presented even though the Tier system is already in place; this is something upon which Council and/or the Plan Team may like to provide input.

Also, either as appendices, or in a separate technical document, it would be helpful to have a number of things. 1) A full model description, including a description of the "dials" and their effects. 2) The ADM code listing. 3) A full set of assumptions and all dial settings for the different models presented. 4) For the final preferred model, a graphical comparison of priors, MCMC posteriors and the approximate posteriors from ADM. For the MCMC, it would also be useful to have diagnostic (e.g. burn in) output.

Presentation interacts with the time-scale upon which the work is carried out and feeds into the decision-making process, as well as with the general "philosophy" of assessment prevalent within a region and the needs for peer review and verification. There are advantages and disadvantages of standardised or bespoke software and models. The variations between assessment methods, and data sources, in the Alaska region make the bespoke model approach sensible. It generally should lead to higher quality assessments and advice, if it is well implemented by good practitioners – and there is no question that this is the case here. Nevertheless, there is a need for verification of code and full documentation of models and assumptions.

In suggesting the two categories of documents above, I envisage a clear, concise annual summary to feed in to decision-making, and a fuller, in-depth set of documents that can be used as a basis for formal peer review and for verification. This second set of documentation should stand for longer than a single year. It is worth noting that verification is far from trivial – either of the general codes used, or of specific codes tuned for annual use. The timing of advisory needs clearly adds to the difficulties of providing formal materials.

Simulation/estimation routines have been informally used for verification and sensitivity testing (of advice, to assumptions or data), and should perhaps be developed further. For example, the implications for management, or of the assumed "Ricker" stock-recruit curve that is used, might be investigated with regard to fitting stock-recruitment "data" derived from assessments using different time-series. This might be a presented as a preliminary council document as well as forming an interesting scientific study worthy of publication.

MULTISPECIES
(Discussions took place with Pat Livingston)

Relatively low-level stomach sampling has been taking place for many years (since mid 80's). A quarterly Multispecies Virtual Population Analysis (MSVPA) has been used to consider multispecies aspects of the Bering Sea system. The major species, in terms of biomass, have all been included. The major interaction is not between species but within the pollock stock, where older pollock impose substantial predation (cannibalism) mortality on younger pollock. Natural mortality at age for juvenile pollock is probably higher than that used in the main assessment – but it also varies year on year due to cannibalism. The decreasing biomass of older pollock indicates a concomitant decrease in the natural mortality of younger fish in more recent years. Clearly, incorporating this feature would have an effect – especially on the perceived stock-recruit scatter, and hence on estimates of F_{MSY} – the quantity of most relevance to management. A pre-specified natural mortality vector could be implemented, or a simple model linking juvenile natural mortality to adult biomass. It is unclear, however, what this would mean in terms of the equilibrium indicators used to guide management, and in management terms it would be "safer" to maintain a constant vector of natural mortality and let the stock-recruit scatter pick up the induced noise due to model mis-specification. This raises a point of general interest - introducing apparent realism is likely to reduce variance but also lead to unknown biases. In terms of a precautionary approach to management (rather than "best biology"), and the use of lower confidence intervals/percentiles of distributions, it is probably better to allow "reasonable" model mis-specification. In terms of pollock management, it is essential to consider broader multispecies issues for some management issues, but not for stock categorisation (under the Tier system) or for calculating catches.

STELLER SEA LIONS/EFFECTS ON STOCK FORECASTING
(Discussions took place with Lowell Fritz)

The rationale for adoption of agreed RPAs (designation of critical habitat and protected areas) is clearly not that pollock fisheries *per se* have caused the observed decline in sea lion numbers. Rather, the rationale is that a decline has occurred, and is continuing, and that the creation of restricted areas for pollock fisheries is a precautionary measure that may enhance the potential of sea lion populations not to decrease further. Pollock constitute approximately sixty per cent of Steller sea lion diet; the sea lion population in the EBS is approx. 25,000; each animal consumes, say, two to three tonnes of fish *per annum* – this means that the total sea lion pollock consumption is thirty to forty five thousand tonnes *per annum*. The annual catch, excluding the international fisheries in the NBS, is of the order of one million tonnes. The issue, at least potentially, is local availability of pollock, especially in the pupping season and at weaning time– July through the autumn.

At face value, there is little of importance for the assessment. However, if inshore boats are forced from prime spawning areas, and with regulations that spread the fishery more throughout the year, it is likely that the exploitation pattern will be

changed. Specifically, it is likely that more juvenile fish will be caught. This has some important consequences.

First, in the 2000 assessment, it will be essential to estimate the selectivity at age for the 1999 fishery and use this in the 2000 (intermediate year) and later projection years. It is difficult to predict how the selection pattern might change in 1999 and later years, and hard to foresee the quality of the selectivity estimates. It may be possible to compare the estimated 1999 fishery selectivity estimates with 1996-1998 estimates to estimate the change in selectivity induced by the pollock fishery restrictions. Only if the change is insignificant, will it be possible to project forward on the basis of an averaged selectivity pattern. If the 1999 only pattern is used, as is likely, additional error will be introduced into relevant parameter estimates and the forecast ABC will be reduced. Attention will need to be paid to this feature in 2000.

Second, the current catch-age matrix shows only small numbers of juvenile fish. If the exploitation pattern shifts towards younger fish, the short-term effect on the assessment will be to introduce an apparent increase in recruits. Because biomass is currently mid-range, this would probably not have much effect on the perceived stock-recruitment relationship – but in principle it could lead to a more peaked relationship – or on the estimation of FMSY. Nevertheless, IF the exploitation pattern shifts towards juveniles, some sensitivity testing will be necessary.

MANAGEMENT
(Discussions took place with Grant Thompson)

Pollock are currently designated as Tier 1, but the ABC is calculated using the Tier 3 formulae. The rationale given for this in the stock assessment and fishery evaluation (SAFE) report is not wholly convincing. Despite all foregoing comments, the assessment-generated pdf of FMSY is arguably “as good as it gets”. The SAFE report says that there is model error and there is a need to be careful. But there is ALWAYS model error and at some point there is a need to make a judgement as to the quality and utility of the assessment-generated quantities of interest.

The Russian data and relationship of the NBS, however, is a point of concern. Also, it is unclear for 2000 and beyond what will happen to the exploitation pattern, given the sea lion RPAs. If there is a move towards juveniles, and the fishery is more spread out in time, the ABC could mean a greater number of fish killed than “intended”. Therefore, the Tier 3 usage may be wise. It is also in line with last year and not problematic in terms of quota which is still increased. Perhaps the most sensible approach at present would be to maintain Tier 1 designation and tier 3 usage until these points have been clarified.

OTHER COMMENTS

Review procedure

The question of the adequacy of the current review procedure was considered. It is difficult to comment on this without also commenting on this review – something that I am hesitant to do. In the case of EBS pollock, the process is less important than the practitioners. It so happens, that at present, the senior assessment scientist, Plan Team, and all collaborators, are of the highest calibre. The annual, internal review is excellent as far as it goes, but the primary purpose is to facilitate the annual round of science, scrutiny and decision-making. My impression is that a regular (possibly triennial) and fuller review of the science and assessment work needs to be accommodated. From experience, I would note that a group of three external reviewers is usually sufficient to provide a range of expertise – and credibility – whilst not being too unwieldy. In addition, to make clear the high standard of the assessments carried out, the senior assessment scientist needs more time to attend to wider scrutiny of the assessment, fuller documentation and verification.